6 Lecture 6

6.1 Transpose of a linear transformation continued

Proposition 6.1. *Let $T, S : V \to W$ be linear transformations. Then $(T+S)^t = T^t + S^t$ $(cT)^t = cT^t$.*

Proposition 6.2. *Let $T : V \to W$ and $S : W \to U$ be linear transformations. Then*

$$(ST)^t = T^t S^t.$$

Proof. Let $\varphi \in U^*$ and $v \in V$. Then $(ST)^t(\varphi)(v) = \varphi(ST(v)) = \varphi(S(T(v))) = S^t(\varphi)(T(v)) = T^t(S^t(\varphi))(v) = (T^tS^t)(\varphi)(v)$. Hence $(ST)^t(\varphi) = (T^tS^t)(\varphi)$ for any $\varphi \in U^*$. We get $(ST)^t = T^tS^t$. ☐

Proposition 6.3. *Let $P = [\mathrm{id}]_\beta^\gamma$ be the change of basis matrix from γ to β. Then the change of basis matrix from β^* to γ^* is P^t. And we have for any $\varphi \in V^*$, $[\varphi]_{\beta^*} = P^t[\varphi]_{\gamma^*}$.*

Proof. By Proposition 5.30, we have $[\mathrm{id}]_{\gamma^*}^{\beta^*} = [\mathrm{id}^t]_{\gamma^*}^{\beta^*} = ([\mathrm{id}]_\beta^\gamma)^t = P^t$. The rest of the proposition follows from Lemma 5.12. ☐

Remark 6.4. By Lemma 5.12, for $v \in V$ we have $[v]_\gamma = P[v]_\beta$ which is the opposite of the change of basis. Thus vectors are also called *contravariant vectors*. However, for a linear function $\varphi \in V^*$, the coordinate change satisfies $[\varphi]_{\beta^*} = P^t[\varphi]_{\gamma^*}$ which aligns with the change of basis of V. So linear functions are also called *covariant vectors*. Physicists use how components transform under a change of basis to define vectors and covectors (linear functions). This distinction is also important in differential geometry; see, for example, [1].

6.2 Double dual space

The dual space V^* consists of linear functions on V. If $\varphi \in V^*$, then φ maps a vector $v \in V$ to a scalar $\varphi(v) \in F$. The double dual space V^{**} is the dual space of V^*. If $\lambda \in V^{**}$, then λ maps a linear function $\varphi \in V^*$ to a scalar $\lambda(\varphi) \in F$.

Example 6.5. Let $V = F^3$ and $\psi_1, \psi_2, \psi_3 \in V^*$ be $\psi_1(x_1, x_2, x_3)^t = x_1$, $\psi_2(x_1, x_2, x_3)^t = x_1 + x_2$ and $\psi_3(x_1, x_2, x_3)^t = x_1 + x_2 + x_3$. We define $\lambda \in V^{**}$ as $\lambda(\varphi) = \varphi(1, 1, 1)^t$ for any $\varphi \in V^*$ which is the evaluation of a linear function at the vector $(1, 1, 1)^t$. Then $\lambda(\psi_1) = \psi_1(1, 1, 1)^t = 1$, $\lambda(\psi_2) = \psi_2(1, 1, 1)^t = 2$ and $\lambda(\psi_3) = \psi_3(1, 1, 1)^t = 3$. This λ is one element in V^{**}. In fact for any $v \in V$, we can define $J(v) \in V^{**}$ by $J(v)(\varphi) = \varphi(v)$ which is the evaluation of φ at $v \in V$. For example $J(0)$ is the zero vector in V^{**} since $J(0)(\varphi) = \varphi(0) = 0$ by linearity of φ. This is true for any $\varphi \in V^*$ and hence $J(0)$ is the zero function on V^* i.e. the zero vector in V^{**}. Theorem 6.7 shows that $J(v)$ for some $v \in V$ are all possible form of vectors that can appear in V^{**}.

Lemma 6.6. *Let V be a finite-dimensional vector space over a field F. Then $\dim V = \dim V^* = \dim V^{**}$.*

Proof. We apply the fact $\dim W = \dim W^*$ to $W = V^*$ and get $\dim V^* = \dim V^{**}$. Since $\dim V = \dim V^*$, we have the lemma. $\square$

Theorem 6.7 (Double dual isomorphism)**.** *Let V be a finite-dimensional vector space over a field F. Define the evaluation map $J : V \to V^{**}$, $J(v)(\varphi) = \varphi(v)$ for $\varphi \in V^*$. Then J is a linear isomorphism.*

Proof. First we show that J is well-defined, that is we show that $J(v) \in V^{**}$ for any $v \in V$. We need to show that $J(v)$ is a linear function on V^*. Let $\phi, \psi \in V^*$, $c \in F$. Then $J(v)(\varphi + \psi) = (\varphi + \psi)(v) = \varphi(v) + \psi(v) = J(v)(\varphi) + J(v)(\psi)$. And $J(v)(c\varphi) = c\varphi(v) = cJ(v)(\varphi)$.

Next we check linearity of J. For $v, w \in V$, $c \in F$, and $\varphi \in V^*$ we have $J(v + w)(\varphi) = \varphi(v + w) = \varphi(v) + \varphi(w) = J(v)(\varphi) + J(w)(\varphi) = (J(v) + J(w))(\varphi)$. In the last equality, we used the definition of the sum of two linear functions on V^*, $J(v)$ and $J(w)$. Thus $J(v+w) = J(v)+J(w)$. And $J(cv)(\varphi) = \varphi(cv) = c\varphi(v) = cJ(v)(\varphi)$. Hence $J(cv) = cJ(v)$. Therefore, J is linear.

Next we show injectivity of J. Suppose $J(v) = 0$. Then for every $\varphi \in V^*$, $J(v)(\varphi) = \varphi(v) = 0$. If $v \neq 0$, we can extend v to be a basis $v, v_2, \ldots, v_n$ and let $\varphi = \varphi_1$ be the first linear function in the dual basis $(\varphi_1, \ldots, \varphi_n)$. Then $\varphi(v) = 1$ while $\varphi(v_i) = 0$ for $i \geq 2$. Thus we have $\varphi \in V^*$ with $\varphi(v) \neq 0$. This is a contradiction. Thus $v = 0$, and J is injective.

Finally by Lemma 6.6 $\dim V = \dim V^{**}$, by Corollary 4.16 an injective linear map $V \to V^{**}$ is automatically surjective. Therefore J is an isomorphism. $\square$

Remark 6.8. If V is infinite dimensional, then J is in general an injective map and not necessarily an isomorphism. When it is an isomorphism, V is called *reflexive* and such spaces are important in Functional Analysis cf. [2].

Remark 6.9. Unlike the coordinates isomorphism ϕ_β which requires a choice of a basis of V, the isomorphism J is naturally defined out of the structure of dual spaces and is independent of a choice of a basis V. Thus we call J the *natural isomorphism* between V and V^{**} and we regard $V = V^{**}$. In category theory, the idea of being natural can be mathematically defined and in that language, "J is a component of a natural transformation from the identity functor to the double dual functor." This is not required in this course.

Proposition 6.10. *Let $T : V \to W$ be a linear transformation. Then $T^{tt} \circ J_V = J_W \circ T$. That is*

$$\begin{tikzcd} V \arrow[r, "T"] \arrow[d, "J_V"'] & W \arrow[d, "J_W"] \\ V^{**} \arrow[r, "T^{tt}"'] & W^{**} \end{tikzcd}$$

Thus if we think of J as the natural identification, then $T^{tt} = T$.

Proof. Let $v \in V$ and let $\varphi \in W^*$. We compute both sides applied to $\varphi \in W^*$. For the left-hand side we have $T^{tt}(J_V(v))(\varphi) = J_V(v)(T^t(\varphi)) = T^t(\varphi)(v) = \varphi(T(v))$. For the right-hand side, $J_W(T(v))(\varphi) = \varphi(T(v))$. Thus $J_W(Tv)(\varphi) = T^{tt}(J_V(v))(\varphi)$, for all $\varphi \in W^*$. Hence $J_W(T(v)) = T^{tt}(J_V(v))$ for any $v \in V$. Therefore, $T^{tt} \circ J_V = J_W \circ T$. $\square$

6.3 Annihilators

Definition 6.11. For a subspace $W \subseteq V$, the **annihilator** of W is defined by $W^\perp = \{\varphi \in V^* : \varphi(w) = 0 \text{ for any } w \in W\}$ which is the set of all linear functions on V which vanish on W.

Proposition 6.12. *Let $\pi : V \to V/W$ be the canonical projection. Then $\pi^t : (V/W)^* \to V^*$ is injective and* $\operatorname{im} \pi^t = W^\perp$. *In particular, $W^\perp \cong (V/W)^*$ and* $\dim W^\perp = \dim V - \dim W$.

Proof. Suppose $\pi^t(\varphi) = 0$ for some $\varphi \in (V/W)^*$. Then for all $x \in V$, $0 = \pi^t(\bar{\varphi})(x) = \varphi(\pi(x)) = \varphi(x + W)$. Hence, $\bar{\varphi} = 0$. Thus $\ker(\pi^t) = \{0\}$ and π^t is injective.

Let $\varphi \in (V/W)^*$. Then for any $w \in W$, $\pi^t(\varphi)(w) = \varphi(w + W) = \varphi(0 + W) = 0$. Therefore, $\pi^t(\varphi) \in W^\perp$, and so $\operatorname{im}(\pi^t) \subset W^\perp$.

Let $\varphi \in V^*$ such that $\varphi \in W^\perp$, i.e., $\varphi(w) = 0$ for all $w \in W$. Define a functional $\tilde{\varphi} : V/W \to F$ by $\tilde{\varphi}(x + W) = \varphi(x)$. We must check that $\tilde{\varphi}$ is well-defined: If $x + W = x' + W$, then $x - x' \in W$, so $x = x' + w$ for some $w \in W$. Then $\tilde{\varphi}(x + W) = \varphi(x) = \varphi(x') + \varphi(w) = \varphi(x') + 0 = \varphi(x') = \tilde{\varphi}(x' + W)$. Thus, $\tilde{\varphi}$ is well-defined.

Linearity of $\tilde{\varphi}$ follows from the linearity of φ.

Now, for all $x \in V$, $\pi^t(\tilde{\varphi})(x) = \tilde{\varphi}(x + W) = \varphi(x)$. Hence, $\varphi = \pi^t(\tilde{\varphi})$. Thus so every element of $W^\perp$ lies in the image of π^t, and therefore $\operatorname{im}(\pi^t) = W^\perp$.

By the first isomorphism theorem Theorem 4.13, $\pi^t : (V/W)^* \to W^\perp$ is an isomorphism. So $(V/W)^* \cong W^\perp$ and $\dim W^\perp = \dim (V/W)^* = \dim V/W = \dim V - \dim W$. $\square$

Corollary 6.13. $W^{\perp\perp} = J(W) = W$

Proof. For any $\varphi \in W^\perp$ and $v \in W$, we have $J(v)(\varphi) = \varphi(v) = 0$. Thus $J(W) \subset W^{\perp\perp}$. By Proposition 6.12, $\dim W^{\perp\perp} = \dim V^* - \dim W^\perp = \dim V - (\dim V - \dim W) = \dim W$. Since J is an isomorphism, $\dim J(W) = \dim W = \dim W^{\perp\perp}$. Thus by Lemma 2.17, $J(W) = W^{\perp\perp}$. $\square$

Remark 6.14. There is a intuitive way to think of the dual space as the mirror of V. The fact that $V^{**} = V$ means if you take the mirror twice, you get back to yourself. If you have a subspace $W \subset V$, then its mirror is W^* (strictly speaking it consists of $\varphi \in W^*$ extended by 0 to all of V, see Homework 6). And $W^\perp$ is the complement of W^* in V^*. So taking $W^\perp$ is taking the complement of the mirror of W. The fact that $W^{\perp\perp} = W$ is saying that if you take complement and mirror twice then you get back to W.

Theorem 6.15. *Let $T : V \to W$ be a linear transformation. Then*

(i) $\ker T^t = (\operatorname{im} T)^\perp$

(ii) $\ker T = (\operatorname{im} T^t)^\perp$

(iii) $\operatorname{im} T^t = (\ker T)^\perp$

(iv) $\operatorname{im} T = (\ker T^t)^\perp$

Remark 6.16. In particular, if $T : V \to V$ is a linear transformation, then T is injective if and only if T^t is surjective. T is surjective if and only if T^t is injective.

Proof. Since for a subspace $W \subset V$ we have $W^{\perp\perp} = W$, (i) and (iv) are equivalent by taking $\perp$. Similarly, (ii) and (iii) are equivalent. Moreover (ii) is exactly (i) applied to the operator T^t using $T^{tt} = T$ from Proposition 6.10. So we only need to prove (i).

Let $\varphi \in (\operatorname{im} T)^\perp$. Then $\varphi(T(v)) = 0$ for all $v \in V$. Since $T^t(\varphi)(v) = \varphi(T(v)) = 0$, we have $T^t(\varphi) = 0$ i.e. $\varphi \in \ker(T^t)$. Thus $(\operatorname{im} T)^\perp \subset \ker T^t$.

Let $\varphi \in \ker(T^t)$. Then $T^t(\varphi) = 0$. That is for any $v \in V$, $0 = T^t(\varphi)(v) = \varphi(T(v))$. Hence $\varphi \in (\operatorname{im} T)^\perp$. Thus $\ker T^t \subset (\operatorname{im} T)^\perp$. □

Corollary 6.17. *Let $T : V \to W$ be a linear transformation and $T^t : W^* \to V^*$ be its transpose. Then we have* $\operatorname{rank} T^t = \operatorname{rank} T$. *In matrix language, we have* $\operatorname{rank} A^t = \operatorname{rank} A$ *for any $A \in M_{m\times n}(F)$.*

Proof. $\operatorname{rank}(T^t) = \dim \operatorname{im} T^t = \dim(\ker T)^\perp = \dim V^* - \dim \ker T = \dim V - (\dim V - \dim \operatorname{im} T) = \dim \operatorname{im} T = \operatorname{rank} T$. □

Remark 6.18. We define $\operatorname{rank} A$ as $\dim \operatorname{im} L_A$ which is the dimension of the span of columns of A and thus this is also called the column rank of A. $\operatorname{rank} A^t$ is the dimension of the span of rows of A and hence named row rank. The last conclusion the the corollary is also referred to as "row rank = column rank".

6.4 Linear systems and Gaussian elimination

Definition 6.19. The **preimage** of a set $Y \subset W$ under T to be $T^{-1}(Y) = \{x \in V : T(x) \in Y\}$. For example $\ker T = T^{-1}(0)$.

A system of linear equation is of the form

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \quad \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m \end{cases}$$

In matrix notation, the equation above is written as $Ax = b$ where $A \in M_{m\times n}(F)$, $x \in F^n$ and $b \in F^m$. In the language of preimage, finding solutions to $Ax = b$ is finding the preimage $L_A^{-1}(b)$.

Lemma 6.20. *For any $y \in \operatorname{im} T$, either $T^{-1}(y) = \emptyset$ or there is $x \in V$ such that $T(x) = y$ and $T^{-1}(y) = x + \ker T$.*

Proof. If $y \notin \operatorname{im} T$, then $T^{-1}(y) = \emptyset$.

If $y \in \operatorname{im} T$, then there is $x \in V$ such that $T(x) = y$. For any $x' \in V$ such that $T(x') = y$, we have $T(x' - x) = T(x') - T(x) = y - y = 0$. Thus $x' \in x + \ker T$. On the other hand, if $x' \in x + \ker T$, then $x' - x \in \ker T$ and $T(x') = T(x) + T(x' - x) = y + 0 = y$. Thus $T^{-1}(y) = x + \ker T$. □

This gives the structure for the preimage $L_A^{-1}(b)$ being either $\emptyset$ or $x + \ker A$ for some $x \in F^n$ such that $Ax = b$. We would like to to develop an algorithm to explicitly find this preimage. So we would need to consider the following problems.

1. Given $A \in M_{m\times n}(F)$, find a basis for $\ker A$, i.e., all solutions to $Ax = 0$.
2. Determine if $b \in \operatorname{im} A$ or $Ax = b$ has a solution.
3. If a solution exists, find a specific x such that $Ax = b$.

The starting point of the algorithm is the following simple lemma.

Lemma 6.21. *Let $P \in M_{m\times m}(F)$ be invertible. Then $Ax = 0$ if and only if $PAx = 0$ and $Ax = b$ if and only if $PAx = Pb$.*

Definition 6.22. An **elementary row operation** on a matrix A is one of the following operations, each corresponding to multiplying A on the left by an appropriate **elementary matrix**.

(i) Swap rows i and j of A: $A \mapsto P_{ij}A$

$$P_{ij} = \begin{array}{cc} \begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 0 & \cdots & 1 & & \\ & & \vdots & \ddots & \vdots & & \\ & & 1 & \cdots & 0 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix} & \begin{array}{c} \\ \vdots \\ \text{row } i \\ \vdots \\ \text{row } j \\ \vdots \\ \\ \end{array} \\ \begin{array}{ccccc} \cdots & \text{col } i & \cdots & \text{col } j & \cdots \end{array} & \end{array}$$

(ii) Multiply row i by $\lambda \neq 0$: $A \mapsto D_i(\lambda)A$

$$D_i(\lambda) = \begin{array}{cc} \begin{pmatrix} 1 & & & & \\ & \ddots & & & \\ & & \lambda & & \\ & & & \ddots & \\ & & & & 1 \end{pmatrix} & \begin{array}{c} \\ \vdots \\ \text{row } i \\ \vdots \\ \\ \end{array} \\ \begin{array}{ccc} \cdots & \text{col } i & \cdots \end{array} & \end{array}$$

(iii) Replace row i with row $i + \lambda \cdot$ row j: $A \mapsto E_{ij}(\lambda)A$

$$E_{ij}(\lambda) = \begin{array}{cc} \begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & \cdots & \lambda & & \\ & & & \ddots & \vdots & & \\ & & & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix} & \begin{array}{c} \\ \vdots \\ \text{row } i \\ \vdots \\ \text{row } j \\ \vdots \\ \\ \end{array} \\ \begin{array}{ccccc} \cdots & \text{col } i & \cdots & \text{col } j & \cdots \end{array} & \end{array}$$

Remark 6.23. If we multiply the elementary matrix to the right of A then we are performing an elementary column operation:
$A \mapsto AP_{ij}$: swaps column i and column j of A
$A \mapsto AD_i(\lambda)$: multiply a column i by λ
$A \mapsto AE_{ij}(\lambda)$: replace *column* j with column j + $\lambda \cdot$ column i

Definition 6.24. Let A be a matrix. We now describe a systematic procedure for solving linear systems by row reduction, known as **Gaussian elimination**.

1. **Pivot selection:** Start with the leftmost column that contains a nonzero entry. Choose a nonzero entry in this column as the *pivot*. If necessary, interchange rows so that the pivot is at the top of the remaining submatrix.

2. **Normalize pivot:** Multiply the pivot row by a nonzero scalar so that the pivot entry becomes 1.

3. **Eliminate below pivot:** For each row below, subtract a suitable multiple of the pivot row so that all entries below the pivot are 0.

4. **Move right and down:** Restrict to the submatrix obtained by ignoring the pivot row and all columns to its left. Repeat steps (1)–(3) until no further pivots can be chosen.

5. **Eliminate above pivot:** Use row operations to eliminate all nonzero entries above the pivot, we obtain the **row reduced echelon form (RREF)**.

Definition 6.25. A matrix is in **row reduced echelon form (RREF)** if
(i) All zero rows, if any, are at the bottom.
(ii) In each nonzero row, the first nonzero entry (called a **pivot**) appears to the right of the pivot in the row above.
(iii) Each pivot equals 1.
(iv) Each pivot is the only nonzero entry in its column.
The non-pivot columns are called **free variables**.

References

[1] Lee JM. Introduction to smooth manifolds. 2003 Springer New York.

[2] Brezis H. Functional analysis, Sobolev spaces and partial differential equations. New York: Springer; 2011.